1 DATA STATISTICS FOR LANGUAGE MODELING

Table 1 shows the data statistics for three benchmarks for language modeling. PTB-TRAIN contains about 41K sentences, with an average length of 21.2 words. Both Yahoo-Train and Yelp-Train contain 100K sentences with an average length of more than 78 words.

2 DIALOG RESPONSE GENERATION ON DAILYDIALOG

Table 2 shows that the conditional VAE gives the worst results. SeqGAN is better than the conditional VAE. iVAEmi gives better results than simple conditional VAE and SeqGAN. Our model gives better BLEU-P and BLEU-F1 scores than iVAEmi.
This suggests that our model can be beneficial for dialogue response generation. We believe that our model can also work for other conditional text generation tasks including text summarization and data2text. We will consider investigating such settings in future.

References


